

AD-A228 948

(1)

DTIC FILE COPY

AD _____

CONTRACT NO: DAMD17-87-C-7187

TITLE: MOLECULAR BIOLOGY OF STLV-III AND HTLV-IV

PRINCIPAL INVESTIGATOR: Joseph Sodroski, Ph.D.

CONTRACTING ORGANIZATION: Laboratory of Biochemistry
Pharmacology
Dana-Farber Cancer Institute
44 Binney Street
Boston, MA 02115

REPORT DATE: August 10, 1989

DTIC
ELECTED
NOV 20 1990
S B D

TYPE OF REPORT: Annual Report

PREPARED FOR: U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The findings in this report are not to be construed as an
official Department of the Army position unless so designated by
other authorized documents.

90 11 19 224

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE				
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION Dana-Farber Cancer Institute Laboratory of Biochemistry Pharmacology	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION		
6c. ADDRESS (City, State, and ZIP Code) 44 Binney Street Boston, MA 02115	7b. ADDRESS (City, State, and ZIP Code)			
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Medical Research & Development Command	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER Contract No. DAMD17-87-C-7187		
8c. ADDRESS (City, State, and ZIP Code) Fort Detrick Frederick, Maryland 21702-5012	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO. 63105A	PROJECT NO. 3M2 63105H29	TASK NO. AC	WORK UNIT ACCESSION NO. 072
11. TITLE (Include Security Classification) Molecular Biology of STLV-III and HTLV-IV				
12. PERSONAL AUTHOR(S) Joseph Sodroski, Ph.D.				
13a. TYPE OF REPORT Annual Report	13b. TIME COVERED FROM 7/1/88 TO 6/30/89	14. DATE OF REPORT (Year, Month, Day) 1989 August 10	15. PAGE COUNT 17	
16. SUPPLEMENTARY NOTATION				
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Virology; Molecular Biology; Biochemistry; Human Immunodeficiency Virus; HTLV IV; STLV III; Simian AIDS; RAI		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) Molecular characterization of the SIV _{mac} 251 clone that exhibits <u>in vitro</u> replication competence, but poor <u>in vivo</u> replication ability was initiated by complete nucleotide sequence determination. Defects in the <u>vpr</u> , <u>nef</u> and <u>env</u> genes were noted and will be the focus of future efforts to restore <u>in vivo</u> replication competence of this cloned virus.				
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL Mrs. Virginia M. Miller		22b. TELEPHONE (Include Area Code) (301) 663-7325	22c. OFFICE SYMBOL SGRD-RMI-S	

FOREWORD

Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification _____	
By _____	
Distribution/ _____	
Availability Codes	
Avail and/or	
Dist	Special
A-1	

Addition information for Methods section.

The pFLB10 proviral DNA was sequenced according to the following method. Oligonucleotide primers were prepared based on consensus sequences for available SIV isolates provided in the Los Alamos data bank. Approximately 85% of these primers were found to yield readable sequence information from the pFLB10 provirus. An additional set of primers was prepared based on the sequence of cDNA clones derived from SIVmac 251 - infected cells. The latter primers were found to prime readable sequencing reactions using the pFLB10 provirus in nearly 100% of the cases. In total, sequencing reactions with approximately 45 different primers were used to complete the sequence of the entire pFLB10 provirus.

Sequencing reactions were carried out according to standard protocols based upon the method of Sanger *et al.* Briefly, template DNA was prepared by treatment of the pFLB10 plasmid with alkali to denature the DNA, followed by strand separation on denaturing acrylamide gels. The primer was annealed to the single-stranded template DNA by heating to 65°C then slowly cooling the reaction to 35°C. The molar ratio of primer:template for most reactions ranged from 1:1 to 3:1. The annealed primer:template was mixed with the 6mM DTT, labeling mix and sequenase enzyme (Sequenase Kit, USB Corporation) and [α -³⁵S] dATP, then incubated at room temperature for 2-5 minutes. The reactions were terminated by addition of ddATP, ddGTP, ddTTP and ddCTP, incubation at 37°C for 3-5 minutes, and addition of stop solution (Sequenase Kit, USB Corporation). Samples were loaded on simultaneous 4% and 8% acrylamide sequencing gels, or in some cases, on 6% gradient gels. Each sequencing reaction typically yielded from 400 to 550 base pairs of information. Each segment of the genome was sequenced using at least two sets of non-overlapping primers.

Molecular Analysis of an Infectious SIV_{mac}251 Proviral Clone

Abstract

Molecular characterization of the SIV_{mac}251 proviral clone pFLB-10 was carried out. This SIV_{mac}251 isolate exhibits in vitro replication competence, but replicates poorly in vivo and, as a consequence is not pathogenic. Defects in the vpr, nef and env genes were noted that may be responsible for attenuations in in vivo replicative ability.

Introduction

The pFLB-10 provirus was derived from a bacteriophage library made from human T cells infected with SIV_{mac}251. The pFLB-10 provirus generates infectious SIV upon DEAE-dextran transfection of HUT78 and CEMX174 lymphocytes. Virus stocks of greater than 5×10^5 reverse transcriptase units per milliliter were used for inoculation of rhesus macaques and cynomolgus monkeys. None of the animals inoculated with the FLB-10 virus has to date demonstrated a persistent viremia or induction of disease, whereas control animals inoculated the uncloned SIV_{mac}251 virus have developed AIDS-like illness in the same time period. The studies in animals are of one year duration and are ongoing. We conclude that changes in the molecularly cloned FLB-10 virus compromise its ability to allow efficient replication in vivo. Here we report the progress made in molecular analysis of this deficiency in in vivo replication/pathogenesis.

Methods

To understand the molecular characteristics of the FLB-10 virus that might be relevant to the observed decrease in in vivo replication, the infectious provirus was completely sequenced using the Sanger dideoxy technique (1). Using chemically synthesized oligonucleotides as primers and increasing the readable sequence information by using gradient urea gels and ³⁵S label, the nucleotide sequence could be obtained using a minimal number of individual priming reactions.

Results and Discussion

The sequence of the 3' half of the FLB-10 provirus is shown in appendix 1 and is summarized in figure 1. The nucleotide sequence of the FLB-10 clone is generally similar to that of other SIV_{mac}251 isolates and is more similar to those isolates than to SIV's derived from mangabeys or African green monkeys. The major open reading frames in the FLB-10 provirus are as follows:

1. vpx

The vpx protein is not necessary for SIV replication, but appears to stimulate virus replication through an unknown mechanism. The vpx open reading frame is present and a potential initiator methionine is evident at the 5' end of the reading frame. The FLB-10 vpx open reading frame could encode a protein of 113 amino acids, with a proline-rich segment near the carboxyl terminus. The vpx of FLB-10 differs from the consensus SIV sequence at one residue (62) M → K.

2. vpr

The vpr protein is not necessary for HIV-1 gene expression, but can stimulate HIV-1 replication through its effect as a promiscuous trans-activator of gene expression. Most SIV_{mac} isolates have an open reading frame for vpr, with heterogeneity at the 3' end. The BK28 isolate encodes a 98 amino acid vpr product, while MM142 encodes a 102 amino acid vpr protein. The vpr open reading frame of the FLB-10 provirus has obviously lost the potential to encode a protein, since the methionine likely to be utilized for initiation, based on sequence similarity to other SIV_{mac} provirus clones, is followed within eleven codons by a stop codon. Following the stop codon, the vpr open reading frame continues with strong sequence similarity to the vpr sequence of other SIV isolates, so it is likely that correction of the single stop codon will result in a full-length vpr product. This potential vpr product would be 98 amino acids long, similar in size to functional vpr proteins observed in the HIV-1 system. In addition, there are two vpr changes in FLB-10 differing from the consensus SIV sequence (47) I → M and (77) C → S.

3. tat and rev

The FLB-10 tat and rev open reading frames are intact, as would be expected for an infectious proviral clone. Tat is a positive trans-activator of viral gene expression, while rev is a post-transcriptional regulator of structural protein expression. The tat gene of FLB-10 differs from consensus by only two changes (27) A → R and (75) S → C, while the FLB-10 rev gene demonstrates no changes from the SIV consensus sequence.

4. env

The FLB-10 envelope glycoproteins are similar to other SIV_{mac} isolates in that a premature stop codon exists in the transmembrane glycoprotein. It appears that either one of a pair of CAG residues is converted to a TAG amber codon in various SIV_{mac} isolates passaged in human cell lines. The FLB-10 clone has a TAG CAG sequence, similar to that seen in the SIV_{mac} 251 but unlike the other SIV_{mac} isolates. In addition, the FLB-10 has a second premature stop codon that would result in a deletion of 18 carboxy-terminal amino acids.

5. nef

Although premature stop codons in HIV-1 nef genes are common, most SIV molecular clones do not have obvious premature nef stop codons. Heterogeneity occurs in the carboxyl terminus of nef in different SIV isolates (from 248-299 amino acid residues). The FLB-10 nef gene, which otherwise could encode a 263 residue protein, contains at least two defects that render it non-functional. First, the nef initiator methionine codon (ATG) has been mutated to an ATA, encoding isoleucine. The next residue, a glycine, which is important for myristilic acid addition, has been preserved in the FLB-10 provirus. Second, a premature stop codon results in only a 94 amino acid product, even if the initiator methionine were present. This result explains the observation that deletions of the FLB-10 provirus in the nef region did not result in phenotypic differences in viral replication rate or in cytopathic effect.

In summary, the molecular changes that might in part account for the attenuated in vivo replication rate of an SIV_{mac} 251 infectious virus include changes in the vpr and nef genes as well as truncation of the transmembrane envelope glycoprotein.

Plans for the next year include mutagenesis of the FLB-10 provirus to examine the effects on in vitro and in vivo replication of restoring some of the mutated vpr, env, and nef sequences. The reading frames that are intact, like vpx and yif, will be deleted to examine the resulting phenotype.

References

1. Sanger F., S. Nicklen and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. PNAS 74: 5463.

Figure 1. Schematic diagram of the FLB-10 3' half of the provirus. The ypx, ypr, tat, rev, env and nef genes of the FLB-10 SIV_{mac} 251 isolate are shown. The X's represent changes in the FLB-10 sequence that render the open reading frame products either unable to be synthesized or prematurely terminated. The dark boxes represent the two coding exons of the functional tat and rev genes of FLB-10.

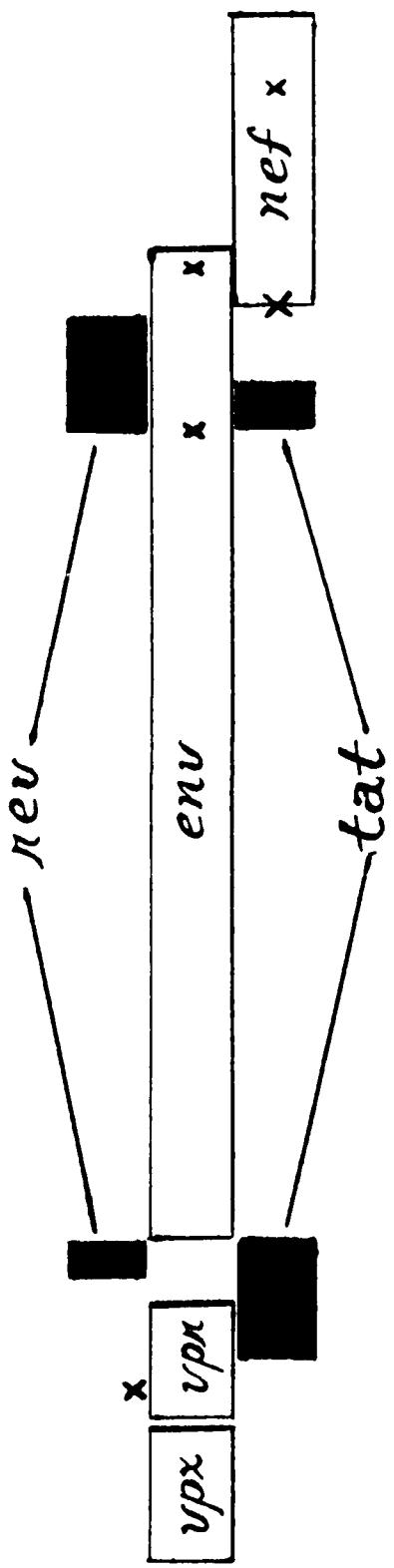


Figure 1.

Appendix 1. Complete nucleotide sequence of the 3' half of the FLB-10 provirus, with open reading frames underlined.

Appendix 1.

SIVmac251 FLB-10 3' sequence

C: 130: 140: 150: (4228 base) (4228) (4228)

1 21
 1: TTTTCAAGGCGGAGAATTC AAGGGCCATCAGGGGAGAAACTGCTGTGCTT
 1: E T A G E I F R A I R G E G I L S C C
 1: . L G P E K G E G P S G E N N C C L A G
 3: . V S G R E E K G H Q G R T T A V L L

61 81 101
 1: TDDCGAGAGCTCATATGATACCAGGTACCAAGCCTACATACITAACACTAAAAGTAGTA
 1: F P F A H I Y Q U P S L Q Y L T L K V U
 3: S R E L I G T R Y Q A Y S T @ H @ K @ P
 3: V P E F S @ V P G T K P T V L N T K S S

131 141 161
 1: AGCGATCTAGATCCCAGGGAGAGAAATCCTACCTGTTAACACTGGAGAGAGAGAAATAAG
 1: S D M R S G G E N P T W M Q W R R D N R
 3: **Vpx** E M S P P P E R I P P G N S G F E T I P
 3: C R C Q I P S F E S H L E T V E R E Q T

181 201 221
 1: AGAGGCCTCGAAATGCTAAACAGAACAGTAGAGGGAGATAAGAGAGAGAGAGAGAGAG
 1: R G L R M A K Q N S R G I K Q R G G K P
 3: S A F E E W L M R T V E E I N R E A V
 3: E R P S N T R T E G @ R P @ T E P @ Q

241 261 281
 1: CCTTAAAGGAGCTTCTTCCAGCTTGGCAAGGGTCTTGGCAAGCTCATATGATACATATG
 1: F T K G A I C E P G L A K V L G I L A E P
 2: L E R E L I F E Q U W Q R S H E Y W L D T
 3: T Y D G S C E S K E G K S L E N T G Y

Vpx
 301 321 341
 1: ACAAAGTAACTCCAGAAATCTTAAATACAGAACACTTGTATGCAAAAAGTTT
 1: C P E U T I C C R I R I L U F N A K S
 3: D I N E A S Y U N Y R Y I C L M P K A
 3: A R C S F K A M Q N T D T C U P C K R

361 381 401
 1: ATTA GCTTCCAGGAACTTCTTGTAGATGTCAGGAACTTAAAGGAGGAGGAGGAGGAGGAG
 1: I Y A L G E F L G M S R G R T R G R G M
 3: F M H C K K G C R C L G E S H G A G G W
 3: Y L C T A S K A U D U E S K D T G Q C P

421 441 461
 1: GAGAGCAAGAGCTCTCTCTCTGCGGAGACTAACATTAATGAGGAAGAAAGCTT
 1: S T P " S S G S P S R Fendix J N S " P
 2: F P G F E I P P P P G L A] S [M (consensus)
 3: T D G T L I E I P L Q P @ H K W R R vpr

~565 stop codon (premature) in vpr
 181 201 221
 1: GAGAGCAAGAGCTC **GAGAG** GAAACCATGGGATEAAATGGCTAGTGGAGGGTTCTGGAG
 1: V K R I T T K G T M G P M S S G S G C R
 3: S A E I F W T E H M I P L L E vpr

581 591 591
 1: ATGAAAGGAAAGCTTAAAGAACATTGATCTGCTGTGCTAAAGTGCATTGCTTCTGCA
 1: I E R R S F K T F @ S S L A N C T W P S
 3: L K S E A L K H E D P E L L T A L S R vpr
 3: N P K K K L P N I L I L A C @ E H L V T

1: Y C S M Y N P H G D T L E S I G E L I R T L O vpr
2: M Y N P H G D T L E S I G E L I R T L O vpr
3: I C I I T M E T P L E S G E N S L E S tat

tat initiation codon

661 681 701
ACGAGCGCTTTCATGCATTTCAGAGGGGATTAACCACTCCAGAATCGGCCACCTGG
1: T S A I H A F G E R I E P L Q N R P T W vpr
2: R A L E M H F R S G S N H S R I G C P G
3: N E R S S C I L E A D R T T P E S A N I tat

721 741 761
GGGAGGAAATCCTCTCACTATAACGGCGCTTGGCGGTGCTATAAACACAIGETATT
1: G R K S S L N Y T A E D L L R R A I T H A I
2: G G N P L S T I P P E G G V L P H M L I vpr
3: G E E I L S Q L Y R P L E A C Y N T C Y tat

781 801 821
GIAAAAAGTGTGCTACCAATTGGCAGTTTGTTCTTAAAGGAACTGGGCGCTTGG
1: V K S V A T I A S F V F L K R D W S env
2: E K V L L P L P V L E S E K G T G I M I
3: C K K D D Y H C Q F C F L K K G L G I tat

Initiation codon rev

841 861 881
~~ATGAGGAGTCAGGAAAGAGAAGAAGAAACTCCGAAAGGCTAAGGCTAATGATTCG~~
1: M C S H E P E E E L R V R L R L I env
2: S A U T K E K K N S E K G P G P F env
3: Y E G S R K S R P T P K K A V A N T C tat

891 921 941
~~CATCAAACAACTAAAGTGGGATGTCCTGGGAATGAGCTGCTTACGCTAT~~
1: H S T E K Y G M S W E S A A Y P F I S env
2: I K D H S M G C L G N S L L I A I env
3: S S P H G V W D V L S T S C L S P E S tat

961 981 1001
AAGTGTCTATGCACTTATGCTACTCAATATGTCAGTGGTCTTATGCTACTGAGTGG
1: K C L W D I L Y S I C L S L L W C T S I
2: S U Y G I Y C T O Y U T V F Y G V P A u env
3: E U S K G S I V L N M S G S F K U Y env

1021 1041 1061
GAGGAATGCGACAAATTCGCTCTCTCTGTCAGCAACGAAAGAGTAGGGATACTTGCGAAAC
1: E E C D N S P L L C N S E G G Y I G A N
2: E N A T I P L E C A T K S R B T W S C env
3: G G M R D F P S S D G P R U S Y L G E R

1081 1101 1121
TCACTGCGTACCGATA TGGTGATIATGAACTGGCGCTTAATGCTTGCGAA
1: S V P T F G W P L F R I G F G C Y F S
2: Q C L P I N G D Y S E I A L N V T E env
3: L S A Y D M V I I G H W P I M A G

1141 1161 1181
TGATGCTTCCAGGAGATGCTGAGAAAGAGAGAGAGAGAGAGAGAGAG
1: C
2: P A L E M T V T E R A T E D U V W R L F E env
3: I M L I P T I Q S T N D E R T Y G N C

1201 1221 1241
GAECTCAATAAGCCTTGCTAAATATCCCATATGCAATTACTATGAGATECAATAA
1: D L N K A L C K I P I M H Y Y E M G G
2: T S I K P C V K L S P L C I T M R C N K env
3: R P Q Q S L V Q N Y P H Y A L L Q D A T

env

1:	K	T	R	S	E	T	B	R	A	G	L	C	T	A	S	T	I	T	N	A	H
2:	<u>K</u>	V	R	O	I	P	G	P	F	G	N	H	O	O	O	O	O	O	O	O	O
3:	Q	H	Q	H	Q	Y	Q	K	K	Q	T	W	S	M	F	L	U	C	U	U	

1321	1341	1361																		
AACATCAGTCACCASTATGACAAAGATACACATGGTCATGAGCTTATTTTGTATAGC																				
1:	N	I	S	T	S	I	R	K	N	F	H	G	Q	P	T	P	C	L	Y	S
2:	T	S	A	P	V	S	E	K	I	D	"	N	E	T	S	T	S	T	S	T
3:	Q	H	Q	H	Q	Y	Q	K	K	Q	T	W	S	M	F	L	U	C	U	

1381	1401	1421																		
TCAGAATAATTGCACAGGGTGGAAACAGGAAATGATAAGCTGGCAATTGACATGCG																				
1:	S	E	G	L	H	R	L	G	T	R	A	N	D	K	L	Q	T	C	H	I
2:	Q	N	N	C	T	G	L	E	Q	E	Q	M	I	S	K	E	N	M	T	
3:	L	R	I	I	A	G	A	W	N	K	S	K	@	@	A	V	N	S	T	R

1441	1461	1481																		
ASGGTTAAAGAGAGACAAAGACAAAGGAACTACAAATGAAAATGGTACTCTACATTTCT																				
1:	R	V	K	K	R	Q	D	K	G	V	G	B	N	L	U	T	Y	R	F	I
2:	G	L	K	E	P	D	K	T	K	E	Y	N	E	T	W	Y	S	J	T	L
3:	R	G	E	K	E	T	R	Q	R	S	T	M	K	L	G	T	T	O	T	W

1501	1521	1541																		
TTGTGAACAAGGGAATAACACTGATAATTGAAAGGAGATGCTGATGATGATGTTACAC																				
1:	L	G	T	R	E	R	H	G	G	P	F	O	M	I	C	N	I	L	S	P
2:	C	E	O	G	N	S	T	D	N	E	S	R	C	Y	M	N	H	C	K	
3:	F	U	N	R	G	I	P	L	I	M	R	A	P	A	T	T	T	T	V	

1561	1581	1601																		
TTCTGTTATCCAAAGAAGTCTTGTCACAAACATTATTGGGATACTATTAGATTAACCTT																				
1:	F	C	Y	P	P	V	L	Q	G	T	L	L	G	Y	Y	S	I	G	J	I
2:	S	U	I	G	E	S	C	R	H	Y	W	P	T	I	H	E	P	Z	I	
3:	L	I	D	S	R	S	L	U	T	N	I	T	G	I	T	E	N	G	I	

1621	1641	1661																
TGCACCTGGAGGTATGCTTTGCTTAGATGATAATGACAAATATTACATTTTATTT																		
1:	C	"	S	F	L	C	F	A	G	"	G	H	F	L	F	"	Y	P
2:	A	F	P	G	Y	A	L	L	F	C	N	E	T	N	"	C	E	F
3:	V	H	L	D	U	M	L	C	L	R	U	M	T	O	I	J	A	

1681	1701	1721																	
TAAGIGTTCTGAGGTGGTGGTCTCTTCATGCACAAAGGATGATGGAGACACAGATTC																			
1:	G	M	E	G	S	G	G	L	F	M	H	K	D	O	G	P	T	R	F
2:	K	C	S	M	V	U	V	S	S	C	I	R	M	M	S	T	S	G	T
3:	L	N	V	L	F	W	W	S	L	H	A	O	G	E	W	S	H	R	L

1741	1761	1781																	
TTGAGTTGGCTTAAATGGAACTGACCATGAGATGAGATTTATGATGATGATGATGATG																			
1:	L	U	W	L	R	W	H	G	S	R	H	R	N	L	Y	A	C	R	F
2:	W	F	G	E	N	G	T	F	A	E	N	T	Y	T	Y	T	A	G	N
3:	I	G	L	A	L	M	E	L	E	R	K	I	E	L	I	F	T	G	M

1801	1821	1841																	
GCAATAACACTATAATTAGTTTAACTTAAGTATTAAATGATGAAATGAAATGAAAG																			
1:	C	A	G	P	V	"	G	A	T	A	M	I	E	N	I	F	"	C	F
2:	S	N	P	T	J	I	I	S	I	A	"	Y	N	L	T	M	E	I	R
3:	G	U	I	G	L	Q	T	L	U	O	T	S	I	I	T	I	T	R	N

1861	1881	1901																		
ACCAGGAAATAAGACAGTITTAACCAAGTCACCATIATGTCTGATGGTGTITCCACTCADA																				
1:	T	R	K	G	D	S	F	T	S	H	H	Y	V	W	I	G	F	P	L	T
2:	P	G	N	K	T	U	L	P	V	T	I	M	S	G	L	U	F	H	S	O
3:	D	Q	E	I	R	O	F	Y	O	S	F	L	C	L	N	W	F	S	T	R

2: P S N R E T
3: N I S M I G G S F I D U G Y E S T I N G S F P

1981 2011 2021
GATAAAGAGGTGAAACAGATAATTCGAAACATCCGAGTAACTGCACTAAATAC
1: N K R G E T P H C O T S Q H Y W N D G
2: T K E V K R T T U K H P R Y T G T N N
3: Q R K E Q N Y F I S N I P G T L E L

env

2041 2051 2091
TGATTAATGAACTTAAAGGGCTCGAGAGAGATCGGAAAGTTACCGTCATGTCAG
1: @ V N Q F N G S G R R S G S Y L H V D
2: D K I N L T A P R G G D P E U T F M W
3: C T K S T @ R L T S E E I F K C P E C

2103 2121 2141
AAGATTGAGAGGAGAGTTCTCTCTCTGATGAAATTGGTTCTAAATTGGTAGAGC
1: K I Q R R U P L L F N E L D S V L G P C
2: N C R G E F L Y D K M N N W E L K N W U E P
3: S T A E E S S E T O K P I S F G T G P F

2151 2161 2201
TTCCATGAACTACCCAGAGGCCAAAGGAAACGGCATAGAAGGAATTACGTGCCCCG
1: E G C N Y P E A K G T A G K E L P A V S
2: R D V T T Q R P K E R H R R N Y U P C H
3: J R H Q L P R S T R N G T E G I T S D

2211 2231 2261
TATG-CACAGAGTAACTCAACATGCTGATAAAGTAGEGCAAAATGTTTATTTGGCTTC
1: Y ? T N N Q H L A Q S R Q K C L P A S
2: I F Q I I N T W H K U G K N U Y L F E
3: J D D K Q S T L S I K G A K R T J C L

2271 2301 2321
AGAGTGAGACCTCACTGTAACCCAGAGCTGACCGTCTCGATAGCAAAATAGTTGAC
1: F G R P H L R L H S D Q S H S K H R L P
2: E G D L E T C N A T U T S I I A N T P W
3: S C E T S R U T F D Q P O T E Q T E I

2341 2361 2381
TGATGAAACGAACTACTGATCAACATGAGTGAGAGCTGGAGAGAGTATTC
1: R W K P N R Y H H E C R G G R T V S I R
2: C G N O T S I T M S A E U A E I Y F I R
3: L M E T K L U S P Q V Q R W D N C J D W

2411 2431 2441
CTTGGAGATTAAAGATTAGTACAGATACTTGGATTGGCTTGCTGACAGATTC
1: C G R L Q T S P P H E T W L G P H F C
2: I G R Y K L Q E I T P I G L A P T Q W
3: C T T T N E T Y S L F L A W F P Q P

2461 2471 2501
GAGGTACACTACTCTGCACTCTCAACAAATAAAAAGAGGGTGTGCTGATGTT
1: E . H Y W W L N K R P E A C A R U
2: C T T T C T T T T T T T T T T T T T T T T
3: G T

2501 2541 2551
GCTTTCTGAAACGAACTACTGCACTGGCTGAGCTGGCTGACGCTGACGCTG
1: G E S F N G R F C N G R D V U D A D R
2: G F L A T A S S A M G A T S L T L T C G
3: W V E S Q R C U L G W A R R R R R P T

2641 2641 2651
 GAGACAAACAAGAATTCTGGGACGAGGGCTCTGGTTCAGAAACTCTAACAACTCT
 : E T T R I V A T B F L E S A C P T T C
 : R O Q E L L R L T V W S T K N I C T T
 : R D N K N C C D G P S E R I S T S F T

2701 2721 2741
 CACTGCCATCGAGAACTTAAGAGGACGGAGACAGCTGGATGTTGGGGATGTCCTW
 : H C H R E V L K G P G T A E C L G M C
 : T A I E K Y L K D Q A G L N A W S S A F
 : S L P S R S T R T R H S R M L G D U R

2761 2781 2801
 TAGACAAAGCTGGCAACATACSTAACTAACATGGCCAAATGAGCTTAAACGAACTGAA
 : P T S L F H Y C T M A K C I S N T E I
 : R O V C H T T V P P R A I S L T E P I
 : L D K S A T C L Y H D M G C P H R T I

2831 2841 2851
 CAAIGATACTTGCGAAAGTGGGAGCGAAAGGTTCCTTTGGGAAAGCTGAACTGAACTGAA
 : Q @ Y L A R V G A K C P L I S C K I S
 : N D T W S E W E R K V I F L E E N I S A
 : T M I L G E S G S E R L T S W R K T E

2881 2901 2921
 CCTCTCTAGAAGAGGACACATTCAACAAAGAGAAAGCTATGAAATTAACTGAACTGAA
 : P P R P G T N S T R E E H V A T T U I
 : L L E E S S I Q Q E K V Y Y E L G C I
 : P S Q R R P R E N K R R T C M N Y T P I

2941 2961 2981
 TAGCTGGGATCTTCTGGATATGGTTGGCTTCTTCTGGGATAAGCTATGAACTGAA
 : @ L G C U W O L U F P C F L D P S V T I
 : S W D V E G N W F E T I A S V I Y T D
 : T A G M C L A T G L T L L C P S T Y I

3001 3021 3041
 TGGAAATTATATAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGT
 : W N L C S C R S N T V K N S D L Y S T N
 : G I Y U U U G U I L L R I U J Y J H D M
 : M E E M G T E E S Y C E P E S T P V K

3061 3081 3101
 GCTAGCTAAAGTAAAGGCAAGGGTATAGGGCACTTCTTCTTCTGGCTACCTCTTCTGG
 : A S P D K A G W P ~~part 2~~ L T P T I L
 : L A H C R P G Y P F K E I S P R S I C
 : C P L S P ~~part 2~~ G I C R G S L P H P I C

3121 3141 3161
 GCAGACTCATACCGAACAGGACCGGCACTTCAACAGAGAAAGGCAAGGAGACGA
 : A D S Y F T G P C T A P Q R R R F R
 : P T C C T P P T L P T S P F G K F P R
 : C R L C A A F T R A L Y R A P P P P P

3181 3201 3221
 TGGAGAAAGGCGGTGGCAACAGCTCTGGCCTGGAGATAGAATATATTCTGGAT
 : W K R R R W Q Q L L A L A D P I Y S F P I
 : G E G G G N S S W P W P I E Y I H F L I
 : V E K A U A T A P G L G R D N I F I S P

env

First premature stop

codon

~~new~~
env
tat

~~new~~
env
tat

3301 3321 3341
 AGCGTACAGATCTCCAAACCAATACTCAGARGGTCCTTGCGAACCCGCGATAT
 1: S I P D P P T N T P E A I C D P T H
 2: R P Q T L D P I L G P T E A T K
 3: E H T R S S N Q Y S P R S L R P Y E C
 Mutated initiation codon nef
 3361 3381 3401
 AGAGWELTCACGACTGAACTGACCTAACAAATAGGTGGACCTATTTCATGKAC
 1: R S P D D Q T D L P T I F V E L E F Q A
 2: E U L R T E L I T Y L O Y R W S Y F H I T
 3: E K S S G L N E E Y N D I G G A I S M D
 env
 env
 nef
 3421 3441 M → I 3461
 AGTCCTGCCGGCTGGAGATCTGCGACAGAAAACCTCTTGCGGGCGGGGGGAGCTTCG
 1: G P S R L E I C D R N S C G R V G S M
 2: U G A G W R S A T E T L A G A W G I I
 3: P S K P A G D L R Q K L Z R A P E
 env
 env
 nef
 3481 3501 Second premature stop codon env
 CGAGAICCTTACGGAGAGGTGGAAGATAATCTGCAATCCCTAGGAGGTTGGCTT
 1: S P S E E R W K I T D P K N P P E R S A
 2: E T L F R G G K R 2nd premature I P R P T I D
 3: G R I L S E V E D R S S C S L G G I P A
 env
 env
 nef
 3541 3561 3581
 GCTTACCTGACCTTCTTGAGGGAGCTGAAATACAGATGGGGGAGGTTGGCTT
 1: A P A H A L U R D P N T I P P S I P
 2: L K L T L L P G T E I D F G A V Y C C C
 3: G I S S R S C E G D K Y N G S G A C
 env
 env
 nef
 3601 3621 3641
 GATGGAGAAATCCAGCTGAAGAAAPGAAAAATTAGATACAGAAAATACAGGCTT
 1: H S D T Q L K K R K K R P T I E N Y I P
 2: M E K P S R P K G K I S T G K T P Y C
 3: P A R A P A E E K E H I A Y P I P P T
 nef
 3661 3681 3701
 AGATATATAGGAAGAATGAGCTTGGTAAGGGCTTAACTGAGGCTTAACTGAGGCTT
 1: I P M R K M M T W Q G Y Q S G S A F P
 2: Y R S G R Q G L G R C I S E A K S S F
 3: I I D E E D D D L V G U S V P K P
 nef
 3721 3741 3761
 GAGCATTGACTTACGAAATTGGAGATACATATGCTTCTTATATAAGATGAGGCTTGCAG
 1: E C P I T N W Q P T G L I D E A S G
 2: S I D L Q I G N I Y U S F Y P F C C
 3: T Q M T Y K L A T D M S H F I M T S
 nef
 3781 3801 3821
 TGGAAAGGATTTTACAGCTGCGAGAGAGATGAAATCTTACACATGAGGCTTACAGGCTT
 1: W K G F I T V Q E D T E S S T C T P K R
 2: C C I I I I I P T T C V I P F T P T C
 3: L E S I I I S A R R P P
 nef
 3841 3861 3881
 AGCAGGCACTACACAGATTTGGAGGATTACCTCTGAGGAGGAACTGATCTG
 1: K K A S Y D I G R I T F D I Q E L C T
 2: R P H H T R L A G L H L R T K N G I P
 3: E X G I I P D W Q D Y T S G P C I F Y
 nef
 nef
 nef

1: E T P I D L W N L P D M A C S C D A D R

3961 3961 premature stop codon ref 400
ATGAGAGGCAATTAAATGGT[]CAGGCTTAAACTCCAGGTGCGAAGGACCCCTGGGGAG
1: M R G I I @ C S Q L K L P S C M T L G E
2: @ E A L F N V ~~W~~ ^W ~~N~~ ^N F Q V G E P L G R
3: D E R H Y L M [E] P A Q T S K W D S P W G

4021 4041 4051
AGGTCTAGGCGGAGAGTITGATTCGAGCTAGCTACACTIATCAGGCATAATGTTAGAT
1: R F P R G S L I Q L P P T L I R H M L D
2: G S S V E V @ S N S S L H L S G I C @ I
3: E V L A W K F D P T L A Y T Y G A Y V R

4061 4101 4131
ACCCAGAAGAGCTTGGAGCAAGTCAGGCTGTCAGAGAGGTTAGAAGAAGGCTAA
1: T Q K S L E A S Q A C Q R K R L E E G F
2: P R R V W K Q V R P V R G P G Q K K A N
3: Y P E S F G S K S G L S E E E V P R R L

4141 4161 4181
CCCCAAGAGGCTTCTAACATGGCTGACAAGAGGAAACTCCCTGAGAGCAGGGACT
1: P Q E A F L T W L T R G K L A E T A G T
2: R K R P S @ H G @ Q E G N S L P Q O G L
3: T A R G L L N M A D K R E T P P D S R D

4201 4221
TTCCAGAAGGGATGTTATGGGGAGGAG
1: E H K G M L W G G .
2: S T P G C Y G E S
3: F P Q G D V M G R ..